

Detecting Risk Drinking during Pregnancy: A Comparison of Four Screening Questionnaires

ABSTRACT

Objectives. This study investigated the efficacy of screening for risk drinking during pregnancy with two brief questionnaires, TWEAK and T-ACE. Both include an assessment of tolerance based on the number of drinks women report they can hold.

Methods. Subjects were disadvantaged African-American obstetric patients in Detroit, Mich. Traditional alcoholism screens (Michigan Alcohol Screening Test [MAST], CAGE) and the tolerance question were administered ($n = 2717$); TWEAK and T-ACE were constructed from tolerance and embedded MAST and CAGE items. In a separate sample ($n = 1420$), only the T-ACE was administered. Periconceptional risk drinking was the gold standard. Screen evaluations were based on receiver-operating characteristic analyses.

Results. At the cutpoint of 2, sensitivity/specificity for embedded screens were 91/77 for TWEAK and 88/79 for T-ACE; comparable values for T-ACE alone were 67/86. TWEAK and T-ACE screened more effectively than CAGE or MAST.

Conclusions. Embedded versions of TWEAK and T-ACE were both highly sensitive to periconceptional risk drinking in this population. Administering T-ACE alone reduced its sensitivity; this suggests that MAST and CAGE administration improves its performance. (*Am J Public Health.* 1996;86:1435-1439)

Marcia Russell, PhD, Susan S. Martier, MSSA, Robert J. Sokol, MD, Pamela Mudar, MPH, Sandra Jacobson, PhD, and Joseph Jacobson, PhD

Introduction

Heavier drinkers tend to ignore general public health warnings to avoid alcohol during pregnancy.¹ This increases the importance of intervention by prenatal care providers to prevent risk drinking among their patients.² Risk drinking during pregnancy has been defined as a level of maternal drinking associated with adverse pregnancy outcome, including developmental deficits in the offspring.³ Because risk drinking during pregnancy is relatively rare, screening patients can reduce the cost involved in the identification and follow-up of risk drinkers.

Two brief alcohol-screening questionnaires, the TWEAK⁴ and the T-ACE,⁵ have been developed and tested on obstetric patients.⁵ The utility of both depends substantially on a question about tolerance to alcohol's effects that assesses alcohol intake indirectly. As tolerance to alcohol develops, increases occur, both in the minimum amount a woman must drink before she begins to first feel its effect, or to feel high, and in the maximum amount she can hold before getting sleepy or passing out. Earlier versions of the TWEAK and T-ACE focused on the minimum, asking, How many drinks does it take to make you feel high?⁵ In the present version, focus is on the maximum, assessing tolerance by asking, How many drinks can you hold?⁶

An indirect approach is preferred because direct questions about how much a patient drinks may trigger denial and minimization of intake, especially in heavy drinkers.⁷ In a recent example of this phenomenon, the sensitivity of a screen for alcoholism fell from 95% to 32% when questions on the quantity and frequency of drinking were asked prior to screening.⁸ Other items included in the screens

were derived from the Michigan Alcoholism Screening Test (MAST)⁹ and the CAGE,¹⁰ two screening questionnaires developed in male populations to detect alcoholism.

The purpose of the present study is to compare the performances of four screens in detecting risk drinking among obstetric patients: the "hold" versions of TWEAK and T-ACE, the MAST, and the CAGE. The evaluation will take into consideration the influence of cutpoints used to define positive scores on the screening questionnaires.

Methods

This investigation is part of an ongoing prospective study initiated in 1987 as part of the Clinical Core of the Fetal Alcohol Research Center in Detroit, Mich. Traditional alcoholism screens, MAST and CAGE, and tolerance items were administered to women on their first visit to a core city prenatal clinic ($n = 3056$) by interviewers trained to elicit alcohol history and consumption information; fewer than 1% ($n = 21$) of the patients declined to participate in the study. Analyses are based on subjects who admitted having drunk alcohol at some time ($n = 2717$); all were African American and had low socioeconomic status.

Marcia Russell and Pamela Mudar are with the Research Institute on Addictions, Buffalo, NY. Susan S. Martier is with the Department of Obstetrics/Gynecology, Hutzel Hospital; Robert J. Sokol is with the School of Medicine; and Sandra Jacobson and Joseph Jacobson are with the Department of Psychology, all at Wayne State University, Detroit, Mich.

Requests for reprints should be sent to Marcia Russell, PhD, Research Institute on Addictions, 1021 Main St, Buffalo, NY 14203.

This paper was accepted March 6, 1996.

TABLE 1—Characteristics of Disadvantaged, African-American Obstetric Patients (n = 2717) in Detroit, by Periconceptional Risk Drinking Status

	Risk Drinking	
	Yes (n = 181)	No (n = 2536)
Maternal age, y, mean \pm SD	28.0 \pm 5.4	24.2 \pm 6.0*
Parity, mean \pm SD	1.9 \pm 1.5	1.2 \pm 1.4*
Gravidity, mean \pm SD	4.7 \pm 2.4	3.2 \pm 2.1*
Prepregnancy weight, lb, mean \pm SD	142.8 \pm 36.2	143.8 \pm 37.7
Gestational age at screening, wks, mean \pm SD	25.4 \pm 9.1	22.6 \pm 8.8*
Smokers, %	86.2	48.2*
Cigarettes/day, mean \pm SD	15.0 \pm 11.5	6.2 \pm 9.1*
Absolute alcohol/day, oz, mean \pm SD	2.6 \pm 2.5	0.1 \pm 0.2*

Note. SD = standard deviation.

*P < .001.

These procedures were approved by an institutional review board for the protection of human subjects. The tolerance item was introduced in June 1988; from April 1989 to October 1990, every other patient was administered T-ACE alone, rather than the MAST, CAGE, and tolerance questions (n = 1420). All other procedures remained constant. To minimize differences in the time periods for comparisons between the T-ACE administered alone and the embedded versions of the TWEAK and T-ACE, current analyses are based on patients seen from June 1988 to October 1990.

Measures

Periconceptional risk drinking. Alcohol consumption was assessed with a procedure suggested by Bowman et al.¹¹ It is comparable to a method currently known as the timeline follow-back procedure.¹² Beginning with a typical weekend (starting Friday afternoon), types of alcoholic beverages (beer, wine, wine coolers, brandy, or other liquor) and the amount of each beverage drunk were recorded for each day of the week. Cognitive interviewing techniques were used to improve reporting.¹³ For example, to help respondents recall their drinking on a typical Friday, interviewers were trained to ask women what they usually did on a Friday afternoon and evening (e.g., did they stay home, go out, have friends over, whatever) and whether they usually drank in that situation. Alcohol consumption data were converted to ounces of absolute alcohol per day.

To assess periconceptional risk drinking, women were asked about their alcohol intake in a typical week before they

became pregnant. Because pregnancy recognition is delayed, prepregnancy consumption was assumed to be representative of consumption around the time of conception. This was confirmed for each woman by calculating the probable date of conception from the date of the last menstrual period and asking if drinking around this date was similar to the reported prepregnancy intake. Periconceptional risk drinking was defined as 1 oz or more of absolute alcohol per day (approximately 14 or more drinks per week).⁷

Tolerance. Women were asked to report the number of drinks they could hold. *Hold* was defined as the number of drinks one could consume before passing out or falling asleep. Few women requested an explanation of this term. An answer of six or more drinks was considered positive, that is, indicative of tolerance to alcohol's effects.

TWEAK. The TWEAK consists of five items⁴ (the italicized words indicate how the acronyms were derived): (1) *Tolerance*—How many drinks can you hold? (2) Does your spouse (or [do your] parents) ever *Worry* or complain about your drinking? (3) Have you ever had a drink first thing in the morning to steady your nerves or get rid of a hangover? (Ever had an *Eyeopener*?) (4) Have you ever awakened the morning after some drinking the night before and found that you could not remember a part of the evening before? (*Amnesia*?) (5) Have you ever felt you ought to *Kut/Cut down* on your drinking? Items 2 and 4 were taken from the MAST, and items 3 and 5 were taken from the CAGE. Positive answers to the tolerance and worry questions score

2 points each, and the last three questions score 1 each, for a possible total of 7 points.

T-ACE. T-ACE consists of four items, three of which it has in common with the TWEAK (tolerance, cut down, and eyeopener).³ In addition, it employs an item from the CAGE: Have people *Annoyed* you by criticizing your drinking? Two points are scored for the tolerance question, and 1 point each is scored for the other three questions, for a possible total of 5 points.

MAST. The MAST consists of 25 questions, many used in previous alcoholism surveys, and was developed to provide a quantitative, structured interview to screen for alcoholism that could be rapidly administered by professional as well as nonprofessional personnel.⁹ MAST items are weighted 0, 1, 2, or 5; when summed, they yield scores ranging from 0 to 53.

CAGE. The CAGE consists of four items: (1) Have you ever felt you ought to *Cut down* on your drinking? (2) Have people *Annoyed* you by criticizing your drinking?; (3) Have you ever felt bad or *Guilty* about your drinking? (4) Have you ever had a drink first thing in the morning to steady your nerves or get rid of a hangover (*Eyeopener*)?¹⁰ Each item receives a score of 1 for a positive response, for a possible total of 4 points.

Procedures

For comparisons of the TWEAK, T-ACE, MAST, and CAGE (n = 2717), it should be noted that the TWEAK and T-ACE were not administered as separate screening instruments; they were constructed from items in the MAST and CAGE and the tolerance question. Data were collected in a routine clinic interview by screeners trained to elicit sensitive alcohol and drug information. Screeners interviewed patients before they were seen by any other health care providers. The order of screening was invariant, with the MAST and the CAGE given first, and then questions on periconceptional drinking and tolerance. This order was adopted in response to pilot studies that suggested that alcohol intakes were underreported when assessed before the MAST and the CAGE. The interview also included questions on maternal and paternal sociodemographic characteristics, maternal reproductive and smoking histories, and paternal alcohol and other drug use. It took approximately 20 minutes to administer all the measures, most of which were

devoted to the MAST and to the ascertainment of periconceptional risk drinking.

Analyses

The following measures of merit were estimated for each screen: sensitivity, the probability that a risk drinker is positive on the test; specificity, the probability that a non-risk drinker is negative on the test; positive predictive value, the probability that a woman with a positive screening score is a risk drinker; and efficiency, the overall percentage of women correctly identified with respect to risk drinking.¹⁴

In addition, receiver-operating characteristic curves¹⁵ for the screens were examined. A receiver-operating characteristic curve provides a representation of performance across a test's entire range of possible cutpoints. The *cutpoint* of a screening test is the value at or above which scores on the test are considered positive. To generate a receiver-operating characteristic curve, sensitivity is plotted on the y axis, and (100 – specificity) is plotted on the x axis for all of a test's possible cutpoints. The ideal screen would be 100% sensitive and 100% specific; this would correspond to the upper left corner of the graph. Accordingly, the cutpoint at which a receiver-operating characteristic curve comes closest to the upper left corner of the graph indicates the cutpoint at which sensitivity is optimized with respect to specificity.

The analysis of the receiver-operating characteristic curves not only allows one to compare different cutpoints for a single screener, but also provides a statistical basis for comparing two or more tests.¹⁶ The area under the curve provides an index of the accuracy of the screening test, or the ability of a test to discriminate between risk and non-risk drinkers, which can be used to compare the efficacy of screeners applied to the same population. This was done using the method described by Hanley and McNeil,¹⁷ in which a z score is calculated that takes into consideration the fact that the areas under the two curves are correlated because they were derived from the same cases. When 95% confidence intervals for the areas under the curves are computed, it is assumed that each curve is based on an independent sample, which is not the case in this study. However, these data are included to permit readers to compare the performance of screens examined in the present study with their performance in other samples.

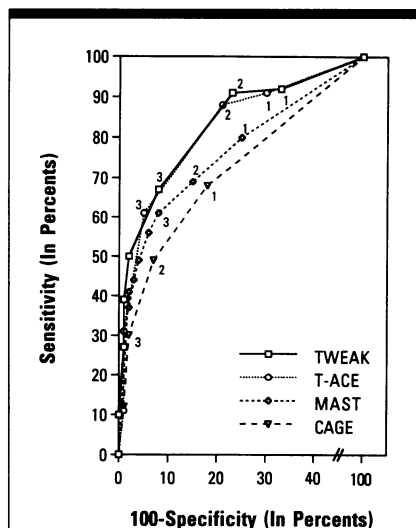


FIGURE 1—Receiver-operating characteristic curves for screening questionnaires predicting periconceptional risk drinking (n = 2717).

Results

Maternal characteristics and substance use data are presented for risk drinkers and non-risk drinkers screened with the MAST, CAGE, TWEAK, and T-ACE (Table 1). Nearly 7% of the women screened were periconceptional risk drinkers. Risk drinkers were significantly older, had given birth and been pregnant more often, and were screened nearly 3 weeks later in their pregnancies than non-risk drinkers. In addition, risk drinkers were more likely to be smokers than non-risk drinkers, and they smoked more heavily.

Receiver-operating characteristic curves for the four screening questionnaires were plotted in Figure 1. The shape of all the curves was comparable; sensitivity increased rapidly as cutpoints decreased, with relatively small decreases in specificity. Accuracy indices, the areas under the curves, are summarized in Table 2. All four instruments were effective in distinguishing risk drinkers from non-risk drinkers. The areas under the curves for TWEAK and T-ACE were similar and significantly larger than those for MAST and the CAGE. By inspection, receiver-operating characteristic curves for the TWEAK and the T-ACE were closest to the upper left corner of the graph at cutpoint 2, indicating optimal combinations of sensitivity and specificity.

TABLE 2—Areas under Receiver-Operating Characteristic Curves (AUC) for Screening Questionnaires (n = 2717)

Screening Questionnaire	AUC (95% CI) ^a
TWEAK	.894 ^b (.867, .921)
T-ACE	.887 ^b (.858, .916)
MAST	.821 (.782, .860)
CAGE	.763 (.720, .806)

^aThese confidence intervals (CIs) assume each AUC is based on an independent sample; they can be used to make meaningful comparisons with AUCs based on other samples.

^bThis AUC is significantly larger than those for MAST and CAGE, based on z scores that take into consideration correlations between the areas because they were based on the same cases.¹⁷

To facilitate comparison of the screens, measures of merit were summarized for cutpoints 1 through 3 in Table 3. At every cutpoint, the 5-item TWEAK was somewhat more sensitive than the 4-item T-ACE. At cutpoints 1 and 2, TWEAK and T-ACE were more sensitive to risk drinking than the 25-item MAST or 4-item CAGE. However, at cutpoint 3, the sensitivity of MAST was comparable to that of TWEAK or T-ACE. CAGE was not particularly sensitive to risk drinking at any cutpoint. At cutpoint 2, only about 22% to 32% of patients who screened positive were found to be risk drinkers, whereas 78% to 90% of the patients were correctly identified by the brief screens.

In the sample of 1420 patients screened with T-ACE alone, values of mean age, parity, gravidity, prepregnancy weight, and smoking were very similar to those observed in the sample of 2717 patients screened with MAST, CAGE, and tolerance questions. However, alcohol consumption was significantly higher in the sample screened with T-ACE alone, 0.4 ± 1.3 oz of absolute alcohol per day compared with 0.2 ± 0.8 oz, and 9.1% of the population reporting risk drinking compared with 6.5%. Sensitivity of T-ACE alone was 67%, specificity was 86%, positive predictive value was 33%, and efficiency was 85%.

Discussion

A key finding was that the two brief screening instruments that included the "hold" version of a question on tolerance

TABLE 3—Measures of Merit (in Percentages) for Screening Questionnaires (n = 2717) at Cutpoints from 1 to 3

Screening Test	Sensitivity	Specificity	PPV	Efficiency
Cutpoint 1				
TWEAK	92	67	17	69
T-ACE	91	70	18	71
MAST	80	73	17	73
CAGE	66	81	20	80
Cutpoint 2				
TWEAK	91	77	22	78
T-ACE	88	79	23	80
MAST	69	84	23	83
CAGE	46	93	32	90
Cutpoint 3				
TWEAK	67	92	39	91
T-ACE	61	95	47	93
MAST	61	91	32	89
CAGE	27	99	56	94

Note. PPV = positive predictive value.

to alcohol's effects, the TWEAK and the T-ACE, were highly sensitive in the detection of periconceptional risk drinking. This finding compares with results obtained in previous research with the "high" version of the tolerance question (the number of drinks it takes to make one feel high), in which sensitivity was found to be lower, 79% for TWEAK and 70% for T-ACE.⁵

It is also important to note that the sensitivity of T-ACE decreased when it was administered alone rather than as part of an interview that included the MAST and CAGE. Although this loss of sensitivity is accompanied by an increase in specificity, sensitivity is generally given priority when screening.¹⁴ Additional research is needed to determine whether the reduction of sensitivity in T-ACE when it is administered alone is reliable, whether the sensitivity of TWEAK is similarly reduced when it is administered alone, and, if so, how MAST and CAGE may condition patients' responses to the TWEAK and T-ACE screening items.

Although current findings suggest that sensitivity of TWEAK and T-ACE is optimized by administering them in the context of MAST and CAGE, the length of MAST makes it cumbersome for clinical use.⁴ Therefore, if MAST indeed seems to be conditioning responses to the screening items, it would be useful to investigate whether this effect can be achieved without administering all 25 questions. MAST includes questions on

severe consequences of drinking, such as hospitalization and legal problems. Mention of more serious consequences of drinking may have made risk drinkers more willing to endorse the less severe items that make up T-ACE. This is the same line of reasoning that led to the development of alcohol consumption measures that employ frequency categories of 3 times a day and quantity categories of 12 or more drinks at a time.¹⁸ Compared with these high levels of intake, having 5 or 6 drinks a day doesn't seem like so much, although it would certainly qualify as heavy drinking by most standards.

The performance of the screens depends on the cutpoint at which they are evaluated. By definition, the highest sensitivity levels were obtained with the lowest cutpoint. However, for TWEAK and T-ACE, there was little loss in sensitivity associated with raising cutpoints from 1 to 2. In contrast, there was a significant increase in specificity associated with the higher cutpoint of 2, and this is important in determining the number of positive screens that must be followed up. In the present study, with 181 risk drinkers and 2536 non-risk drinkers, every 1% increment in sensitivity meant that the screener identified 1.8 more risk drinkers, and every 1% decrease in specificity meant that an additional 25.4 non-risk drinkers were positive. Thus, increasing the cutpoint for TWEAK from 1 to 2 decreased sensitivity by only 1% (i.e., 2 fewer risk drinkers were identified), but specificity

increased by 10%, which meant that approximately 254 non-risk drinkers were no longer positive on the screen.

Positive predictive values and efficiency are influenced by the prevalence of the condition for which one is screening.¹⁴ When the prevalence of a condition is low, as it is for periconceptional risk drinking, high positive predictive values and efficiency are associated with screens having high specificity rates. Unfortunately, specificity tends to be inversely related to sensitivity. Currently, it is not possible to screen effectively for pregnancy risk-drinking and still obtain high positive predictive values and high levels of efficiency. Providers of prenatal care must either keep the cost of following up positive screens low enough to permit sensitive screening for risk drinking or accept the fact that they will miss more risk drinkers in their patient population by employing more specific, less sensitive screening procedures.

One limitation to studies of this nature is that an objective measure of alcohol intake that could serve as a "gold standard" is lacking. Despite an intensive search for a reliable, valid biomedical marker,^{19,20} there is consensus that well-designed self-report measures provide the best available method of determining alcohol intake.^{21,22} A number of procedures to enhance the validity of self-report data were employed in the present study. Interviewers were trained to ask about alcohol use in a sensitive manner. It was emphasized to respondents that the time frame for questions on alcohol consumption was *prior* to pregnancy, to avoid any stigma attached to admitting large alcohol intakes during pregnancy. Also, asking patients to reconstruct drinking events by thinking about their preferred beverage, usual companions, and activities associated with drinking, and so forth provided cognitive cues to assist recall. In addition, predictive validity for the present measure of risk drinking was provided by reports that it was significantly related in a dose-response manner to craniofacial abnormalities associated with fetal alcohol syndrome.^{23,24}

Future studies should investigate the utility of self-administered forms of the brief questionnaires in obstetric and gynecologic populations. For example, a self-administered form of the TWEAK has been successfully used in community samples, with hospital clinic outpatients, and with alcoholic inpatients.²⁵ Although the literacy of the populations involved must be taken into consideration, self-

administered questionnaires, whether by paper and pencil or by interactive computer programs, have the potential of reducing the cost of screening. There is also evidence to suggest that women actually report more alcohol use and alcohol problems on self-administered²⁶ or computer-administered²⁷ questionnaires than in face-to-face interviews.

The generalizability of our findings is limited. The program of research conducted in Detroit, Mich,^{3,5,6} indicates that versions of TWEAK and T-ACE based on the "hold" tolerance question are more effective than those based on the "high" tolerance question in screening for risk drinking among disadvantaged African-American obstetric patients attending an inner-city clinic. However, there may be regional and sociodemographic differences in how individuals respond to the alternative tolerance questions. For example, upper- and middle-class women may find it more socially acceptable to say that they can have several drinks before beginning to feel *high* than to acknowledge that they can *hold* a large number of drinks without passing out. Differences in responding may also reflect possible differences in drinking patterns. The "high" question may be more relevant to women who frequently have three or four drinks, but never drink to the point of passing out. In contrast, the "hold" question is sensitive to drinking patterns that include occasions on which large amounts of alcohol are consumed at one time.

Significant damage to the fetus can take place early in pregnancy, before a woman even realizes that she is pregnant; therefore, intervention is most effective if it takes place prior to conception. However, intervention during pregnancy has the potential of reducing growth retardation and minimizing the effect of alcohol exposure on brain development, which takes place throughout pregnancy.²⁸ In addition, factors often associated with pregnancy, a physiologically based aversion to alcohol and increased social support, enhance the effectiveness of intervention.²⁹ Successful intervention in problem drinking can improve a mother's ability to care for her child, and, most importantly, protect subsequent pregnancies.⁴ For all these reasons, screening for

risk drinking during pregnancy is an important part of good prenatal care. These promising results support increased research on the use of TWEAK and T-ACE in obstetric populations, and suggest extending their use to include women at risk of becoming pregnant. □

Acknowledgments

The research was supported by grants R01-AA07606 and R01-AA06966 from the National Institute on Alcohol Abuse and Alcoholism.

References

- Hankin JR. FAS prevention strategies: passive and active measures. *Alcohol Health Res World*. 1994;18:62-66.
- Minor MJ, Van Dort B. Prevention research on the teratogenic effects of alcohol. *Prev Med*. 1982;11:346-359.
- Sokol RJ, Martier SS, Ager JW. The T-ACE questions: practical prenatal detection of risk-drinking. *Am J Obstet Gynecol*. 1989;160:863-870.
- Russell M. New assessment tools for drinking in pregnancy: T-ACE, TWEAK, and others. *Alcohol Health Res World*. 1994;18:55-61.
- Russell M, Martier SS, Sokol RJ, et al. Screening for pregnancy risk-drinking. *Alcohol Clin Exp Res*. 1994;18:1156-1161.
- Martier SS, Sokol RJ, Bottoms S, et al. Optimized screening for risk drinking: tolerance reconsidered. *Alcohol Clin Exp Res*. 1990;14:315. Abstract.
- Sokol RJ, Martier S, Ernhart C. Identification of alcohol abuse in the prenatal clinic. In: Chang NC, Chao HM, eds. *NIAAA Research Monograph 17: Early Identification of Alcohol Abuse*. Washington, DC: US Department of Health and Human Services. 1985;209-227. DHHS publication ADM 85-1258.
- Steinweg DL, Worth H. Alcoholism: the keys to the CAGE. *Am J Med*. 1993;94:520-523.
- Selzer ML. The Michigan Alcoholism Screening Test: the quest for a new diagnostic instrument. *Am J Psychiatry*. 1971;127:1653-1658.
- Ewing JA. Detecting alcoholism: the CAGE questionnaire. *JAMA*. 1984;252:1905-1907.
- Bowman RS, Stein LI, Newton JR. Measurement and interpretation of drinking behavior: I. on measuring patterns of alcohol consumption; II. relationships between drinking behavior and social adjustment in a sample of problem drinkers. *J Stud Alcohol*. 1975;36:1154-1172.
- Sobell LC, Sobell MB. Timeline follow-back: a technique for assessing self-reported alcohol consumption. In: Litten R, Allen J, eds. *Measuring Alcohol Consumption*. Totowa, NJ: Humana Press; 1992:41-72.
- Friedenreich CM. Improving long-term recall in epidemiologic studies. *Epidemiology*. 1994;5:1-4.
- Hennekens CH, Buring JE. Screening. In: Mayrent SL, ed. *Epidemiology in Medicine*. Boston, Mass: Little, Brown & Co; 1987:327-347.
- Metz CE. Basic principles of ROC analysis. *Semin Nucl Med*. 1978;8:283-298.
- Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*. 1982;143:29-36.
- Hanley JA, McNeil BJ. A method of comparing the areas under receiver operating characteristic curves derived from the same cases. *Radiology*. 1983;148:839-843.
- Cahalan D, Cisin IH, Crossley HM. *American Drinking Practices: A National Study of Drinking Behavior and Attitudes*. New Brunswick, NJ: Rutgers Center of Alcohol Studies; 1969.
- Chan AWK. Recent developments in detection and biological indicators of alcoholism. *Drugs Soc*. 1993;8:31-67.
- Rosman AS, Lieber CS. Biological markers of alcoholism. In: Lieber CS, ed. *Medical and Nutritional Complications of Alcoholism: Mechanisms and Management*. New York, NY: Plenum Medical Book Co; 1992:531-563.
- Midanik LT. Validity of self-reported alcohol use: a literature review and assessment. *Br J Addict*. 1988;83:1019-1029.
- Room R. Measuring alcohol consumption in the US: methods and rationales. In: Clark WB, Hilton ME, eds. *Alcohol in America: Drinking Practices and Problems*. Albany, NY: State University of New York Press; 1991:26-50.
- Ernhart CB, Sokol RJ, Martier S, et al. Alcohol teratogenicity in the human: a detailed assessment of specificity, critical period, and threshold. *Am J Obstet Gynecol*. 1987;156:33-39.
- Ernhart CB, Sokol RJ, Ager JW, Morrow-Tlucak M, Martier S. Alcohol-related birth defects: assessing the risk. *Ann N Y Acad Sci*. 1989;562:159-172.
- Chan AWK, Pristach EA, Welte JW, Russell M. Use of the TWEAK test in screening for alcoholism/heavy drinking in three populations. *Alcohol Clin Exp Res*. 1993;17:1188-1192.
- Russell M, Bigler L. Screening for alcohol-related problems in an outpatient obstetric-gynecologic clinic. *Am J Obstet Gynecol*. 1979;134:4-12.
- Lapham SC, Kring MK, Skipper B. Prenatal behavioral risk screening by computer in a health maintenance organization-based prenatal care clinic. *Am J Obstet Gynecol*. 1991;165:506-514.
- Coles C. Critical periods for prenatal alcohol exposure: evidence from animal and human studies. *Alcohol Health Res World*. 1994;18:22-29.
- Rosett HL, Weiner L, Edelin KC. Treatment experience with pregnant problem drinkers. *JAMA*. 1983;249:2029-2033.